Alissa McBain
Wrangle and Analyze Data – WeRateDogs

I recently dug into some data on the WeRateDogs Twitter account. The data included tweets (only those with ratings), dog type (doggo/floofer/puppo/puppers), dog name, and other Twitter information. Additionally, the photos from these tweets were put through a neural network to see if they were actual photos of dogs. The neural network gave three predictions for each photo, and tried to predict, not only if the photo was actually of a dog, but also the breed of dog.

I was curious to see how accurate this neural network was at predicting if the photo was actually a dog or not. Luckily, the photos had been checked by a human to confirm whether the subject of the photo was a dog or not. So how well did the neural network perform?

Each of the three predictions that the network output for each photo was accurate between 73.7% and 75.2% of the time. I not an expert on neural networks, but I'd say that is pretty great! One interesting thing to point out here, the accuracy of the prediction did *not* decline between the first and the third predictions.

Secondly, I was interested in knowing which dog type (doggo/floofer/puppo/puppers) had the highest WeRateDogs rating.

good_dog_type	count	mean	std	min	25%	50%	75%	max
doggo	75.0	11.853	1.430	8.0	11.0	12.0	13.00	14.0
floofer	10.0	11.800	1.033	10.0	11.0	12.0	12.75	13.0
pupper	234.0	10.821	2.032	3.0	10.0	11.0	12.00	27.0
puppo	25.0	12.080	1.256	9.0	12.0	12.0	13.00	14.0

I generated this table of interesting information and discovered that pupper was the most common dog type, but, surprisingly, had the lowest average rating. Puppos had the highest average ratings, and also one of the smallest ranges in ratings. Puppos had seemingly no outliers, as all of their ratings were between 9/10 and 14/10.

Lastly, I was curious to see if a high WeRateDogs rating corresponded to more favorites or retweets.

Rating vs Favorite Count
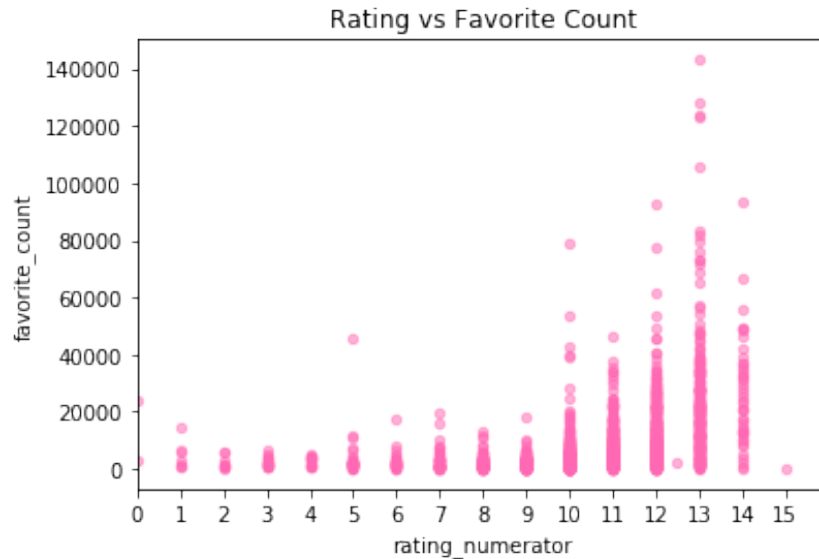
The above graph shows the relationship between the rating and how many times it was favorited. I definitely see a relationship, though not a linear one. The lower rated photos had fewer favorites, while the higher rated photos have a higher chance of being favorited.

Rating vs Retweet Count
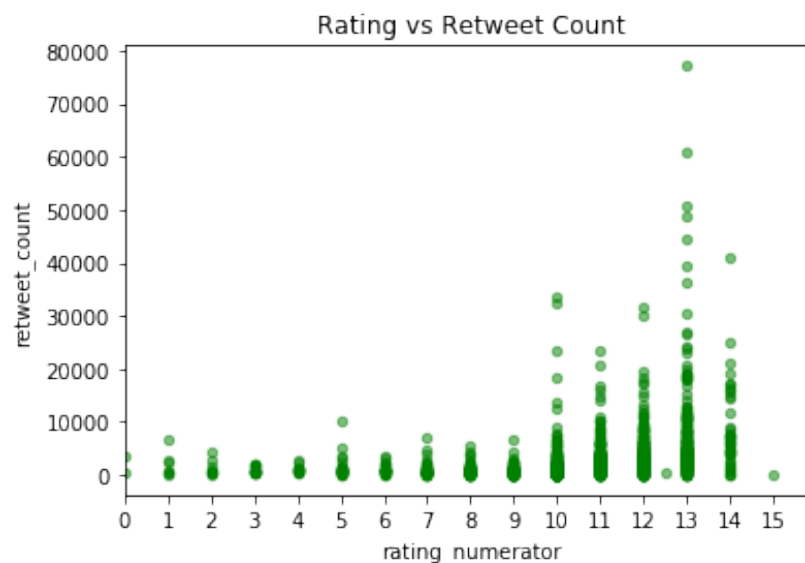
This graph shows a similar relationship, only this time we're looking at rating vs. how many times it was retweeted. Again, the lower the rating, the fewer retweets it received, but the higher rated tweets had a greater likelihood of being retweeted many more times. The relationship is not linear, but with a transformation I'm confident we could make a decent prediction model.